图书情报工作
第63卷 第11期 2019年6月

基于深度学习CNN模型的图像情感特征抽取研究*

李志义[1] 许洪凯[1] 段斌[2]
[1] 华南师范大学经济与管理学院 广州 510006 [2] 华南师范大学信息光电子科技学院 广州 510006

摘要:[目的/意义]以用户情感为线索的图像检索已成为机器学习研究的热点,但图像情感特征标注的语料数据多来源于对图像低层特征的抽取,从而导致图像检索过程单一化和程式化。本文提出了一种基于深度学习的图像情感特征抽取的算法,将图像底层特征融合到图像的高层情感语义当中,为实现图像的情感语义检索提供了参考。[方法/过程]利用改进的卷积网络模型,将数据集图像的颜色、纹理作为输入,经多层运算自动提取图像的情感信息,并通过反向传播算法计算出改进后模型的情感检索准确率,构造出准确率较高且过拟合程度低的图像情感特征提取模型。[结果/结论]应用改进的卷积神经网络模型,实现了对图像情感特征的抽取,相较于原模型提升了10%的检索准确率。

关键词:深度学习 图像 情感特征 抽取 卷积神经网络
分类号:TP391.1
DOI:10.13266/j.issn.0252-3116.2019.11.011

随着微博、各类论坛、知乎、豆瓣等社交媒体的快速发展,越来越多的人开始在社交网络上表达自己的观点、共享知识、分享信息、创作内容。这些网络用户生成海量的内容,往往带有主观情感倾向性,基于文本特征和自然语言处理的情感分析在该阶段发挥了较大作用,并对用户的情感倾向进行了有效地把握。如今,伴随移动社交平台的普遍使用,拍照上传照片表达和传递情感,也已成为大众喜好的信息交流方式。对这些图片加以分析、整理和组织,有效地挖掘这些图像数据中的情感信息,对热点探究、舆情分析具有重要意义[1]。但是,由于图像表达信息的丰富性和人的主观认知因素多样性,造成了基于情感的图像特征抽取具有较大难度。

从情感角度出发,以图像情感语义为线索检索图像,有助于获取用户观赏图像信息时的情感状态,优化用户情感特征库,完善基于情感语义的图像检索系统。目前,图像情感特征标注的语料大多源于对图像低层特征的抽取,使得图像检索过程单一化、程式化。本文试图提出一种基于VGGNet-16模型的图像情感特征抽取的算法,将图像底层特征融合到图像的高层情感语义当中,弥补现有的研究在情感语义特征抽取方面中的不足,以达到更全面、系统地揭示图像特征的目的,为实现图像的情感语义检索提供参考。

1 相关研究综述

针对用户情感特征的研究,以用户生成文本为语料的情感分析方法已较为成熟,通过对文本的情感特征抽取,探索用户情感倾向,在实践应用中已成为探究用户情感线索的重要方式。如马松岳等[2]利用ROSTCM 6.0情感分析工具,对豆瓣网用户关于电影的评论进行了情感抽取,将用户生成的评论文本进行分词,计算出每一个词的情感值并以可视化的形式展现,从而得出用户的情感倾向,为电影附上情感标签;蒋知义[3]等利用自建情感词典,对用户生成的微博内容进行分词并计算情感值,将用户情感倾向划分为正向、负向、中性情感,从而判断用户情感演化趋势。基于文本内容的情感分析其局限在于对非文本模态的数据不具备情感分析能力。随着自媒体的快速发展,图片已成为用户生成信息的主要类型,并且比文本更能体现用户的情感[4],图像情感分析开始进入研究视野。

图像情感语义的抽取,大致经历了两个发展阶段:其一是基于图像视觉特征的情感语义提取阶段;其二

* 本文系国家社会科学基金项目"基于表示学习的跨模态检索模型与特征抽取研究"(编号:17BTQ062)研究成果之一。
作者简介:李志义(ORCID:0000-0001-6407-2554),副教授,硕士,硕士生导师,E-mail:Leeds@scnu.edu.cn;许洪凯(ORCID:0000-0002-3304-4312),硕士研究生;段斌,本科生。
收稿日期:2018-08-08 修回日期:2018-12-19 本文起止页码:96-107 本文责任编辑:杜杏叶

是机器学习阶段,尤以深度学习为代表,其理论广泛应用于图像与视频分析、计算机视觉、语音识别、多媒体等诸多领域,并取得了巨大成功[5]。在此阶段,机器学习方法被应用到图像情感语义特征的抽取。典型的算法模型有:①深度信念网络(Deep Belief Network,DBN)。该模型在图像分类领域取得较大的研究成果,它通过分步训练每一层网络,将图像特征映射到不同特征空间,以概率分布的方式建立观察数据与特征标签的联合分布,提升图像分类的精度[6]。②深度残差网络(Deep Residual Network,DRN)。该模型在计算机视觉、多媒体数据处理等跨模态数据领域引起了广泛的关注,它针对图像特征抽取准确率迅速达到饱和并逐渐趋于下降的"退化问题",在所增加的网络层次上采用了恒等映射的方法,以达到控制训练结果误差的效果[7]。

相对而言,国外在图像抽取方面的研究起步较早,也取得了进展和突破。J. A. Black 等[8]将主题标引和情感标引进行了对比实验,发现两者在聚类结果上呈现出较为显著的一致性,一定程度地反映了以用户情感为线索的图像标引和检索的可行性。K. Yoshida 等[9]提取图像颜色特征,将用户对图像的视觉感受与图像特征建立映射,完成图像情感语义特征的提取;S. B. Cho 等[10]用高兴、沮丧、凉爽来描述图像,建立了支持该类情感的图像情感语义提取系统;C. Colombo 等[11]用愉快、紧张、放松、动感来描述图像的情感语义,建立了图像与情感的映射;S. Siersdorfer 等[12]借助颜色直方图提取图像的颜色特征,利用 SentiWordNet 英文情感词典为图像携带的文本信息打分,提取图片的情感倾向,最后利用 SVM 分类器将颜色特征与情感特征相结合,揭示图像的情感特征。这些图像情感提取方法,都将图像视觉特征作为情感语义抽取的对象,依赖于人对基本情感的分类,受研究者的主观思维所影响。而基于深度学习的图像情感特征抽取可利用算法和模型计算情感值,相对客观地对图像、文本进行了情感归类和分析,近几年已逐渐成为图像情感特征抽取研究的热点。如:Quanzeng You 等[13]利用 CNN 算法,对 Flickr 图像数据集进行情感分类,将图片在情感上分为正负两类,并利用主成分分析法,提出了 RCNN 模型,用于图像的情感特征提取;在此基础上,O. Vinyals 等[14]利用 CNN 模型对图像情感特征进行抽取,将经过人工翻译与描绘的图像内容与图像情感倾向相结合,生成描述图像内容的语句,实现了"看图说话",并提出了图像情感内容自动描述的设想;Manli Sun 等[15]在探究利用 CNN 模型实现图像情感分类时,为减少模型权值训练数量,在池化层插入了权值共享,保持图像情感特征抽取观测样本的稳定性。

我国学者在图像情感特征抽取领域也进行了相关研究。黄崑等[16]在研究中指出,图像情感特征来源于图像、用户与环境,提出了以用户感受为中心描述和刻画图像并进行图像检索的观点;王上飞等[17]提取图像 RGB 的分布特征,利用径向基函数神经网络进行图像情感特征提取,初步形成基于图像内容的情感特征模型;武频等[18]利用支持向量机算法实现小样本图像自动情感注释,设计了基于情感语义的图像检索系统;王华秋等[19]在图像情感语义模型中,将图像的形状矩和颜色矩作为模糊神经网络输入,利用遗传算法对图像的情感映射进行了优化,提高了图像匹配的效率。

国内外对图像情感语义特征抽取的研究,从依赖于图像低层特征的情感特征抽取,逐渐过渡到基于深度学习算法的图像情感语义特征抽取,已取得较大研究进展。但在所建立的特征抽取模型上,仍有巨大的改进空间,以提升算法的优越性。目前,图像情感领域出现了以人脸情感识别为代表的人脸识别技术,这就需要学者在图像情感的基础上研究表情识别,该课题在计算机视觉领域颇具挑战性[20]。人类情感信息的特殊性和处理人类情感的复杂性,决定了将图像情感与人类情感相结合的特征抽取方法仍有巨大的发展空间。

2 图像底层特征抽取的一般方法

要进行图像情感特征抽取,首先要抽取图像的底层特征,通过底层特征联结高层情感语义,建立图像的情感特性。

2.1 图像底层语义特征与情感

图像中最基本的底层可视化内容是颜色、纹理等特征,人们在观察图像时,首先会接收到这些视觉特征,进而受到影响产生不同的情感体会。

在图像视觉特征中,颜色特征是对图像情感最显著的表达,是最具情感特性的特征。针对图像的情感表示,颜色特征能够对应不同的用户主观感受[21]。从多数心理学实验得出的常见情感和颜色相对应的实验结果来看,颜色特征集中的红色,对应了兴奋、热情、奋进、狂怒和强烈;绿色对应了清爽、宁静、自然、疲劳等情感;再细分用户的情感,情感集中包含了正面情感和负面情感,因此一种颜色也关联着不同的情感倾向,进而又可以推论出某些情感可能又关联着不同的颜色[22]。

对图像颜色特征的抽取，通常以RGB颜色空间与HSV颜色空间为依据。从人类视觉感知上看，HSV系统比RGB系统更接近于人们的心理反应和对彩色的感知，从而更接近人眼所观察到的事物的颜色属性，它具有视觉一致性[23]。

通过对图像颜色特征的分析，不难发现颜色是图像特征最显著的部分，同时也与情感语义关联最为密切，因此，如何有效抽取颜色特征是连接图像高层情感语义的关键[24]。

纹理特征也会对人的心理产生一定的作用。虽然它没有颜色特征那么明显和直观，但由于纹理特征中包含了对比度和空间频率等因素，因而不同纹理能够产生不同的视觉效果，进而对人类的情感产生了影响。

2.2 图像颜色特征与纹理特征抽取算法

在图像颜色特征抽取算法中，全局直方图的运用最为广泛，即抽取整个图像的颜色分布表示为该图像的颜色特征[25]。全局颜色直方图描述了不同颜色在同一幅图像中的比例，一般图像的全局直方图实际上是一个1－D离散函数，定义为：

$$H(k)=\frac{n_k}{N} \qquad k=1,2,3,\cdots,L-1 \qquad 式(1)$$

其中，k指的是该图像的灰度取值，L代表灰度可取值的个数，n_k是图像中具有灰度值为k的像素总个数，N为图像像素的总数。

颜色直方图的优点是计算简单，只需计算出图像中每一种颜色的像素频数，且对图像的平移、旋转或者缩放等变化具有很强的稳定性[26]。目前研究人员主要运用Matlab进行图像颜色特征的提取。

图1是基于颜色直方图的图像特征提取的结果：首先分别处理H、S、V三个特征集，获得若干个小的颜色区间；然后将对应的小区间生成颜色直方图的bin（色区），通过计算颜色落在每个小区间的bin绘制得到量化的颜色直方图；得到图2的颜色空间维度量化表，即获取了图像的颜色特征。其缺点是，对于颜色分布较为相似的图像无法准确辨析，如图2中的红色与绿色量化结果相似，没有明显的区别，易造成同色不同情感的图像相互混淆；另一方面，全局颜色直方图无法考虑到图像中各个区域的特征[27]。

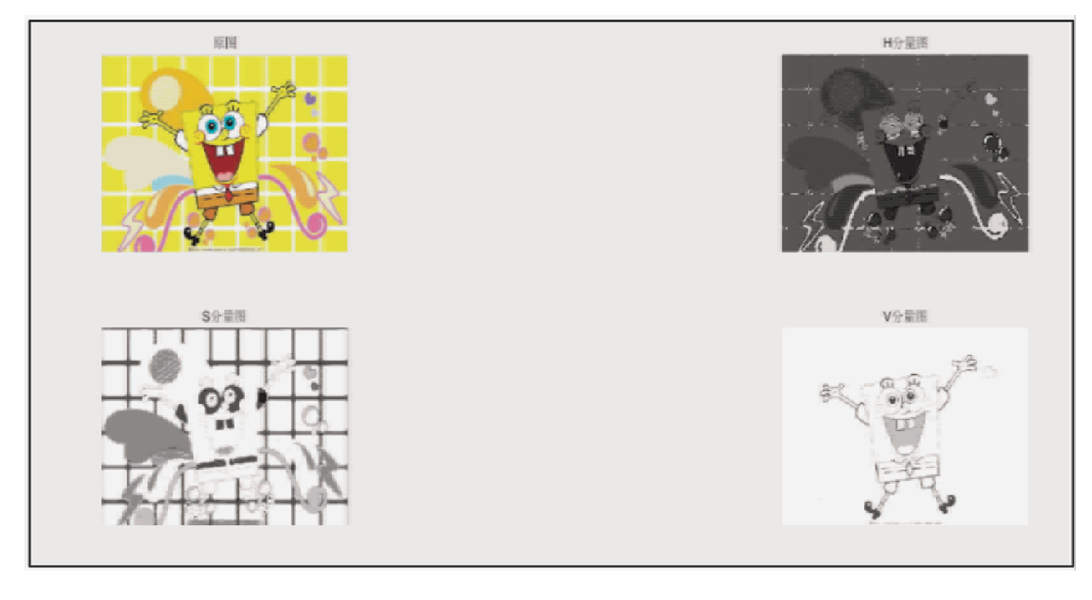


图1 基于颜色直方图的图像特征提取

图像来源：http://pic28.photophoto.cn/20130712/0007020101331274_b.jpg

图像纹理特征的抽取，目前主要采用基于统计的方法和几何法。其中，基于统计方法的灰度共生矩阵（GLCM）法比较常见。

上述传统的图像特征抽取算法能较好地获取图像底层语义特征，却无法准确地表达图像情感，需要借助人的主观判别建立起情感关系，所抽取的情感倾向具有明显的主观倾向性，无法客观地表达图像高层情感语义特征。

因此，本文尝试利用深度学习方法建立图像底层语义到高层情感特征的映射，实现图像情感特征的抽取。

3 基于改进的CNN模型的图像情感特征抽取

深度学习源于人工神经网络的研究，是机器学习的一个新兴分支。其核心在于借助算法使得计算机能

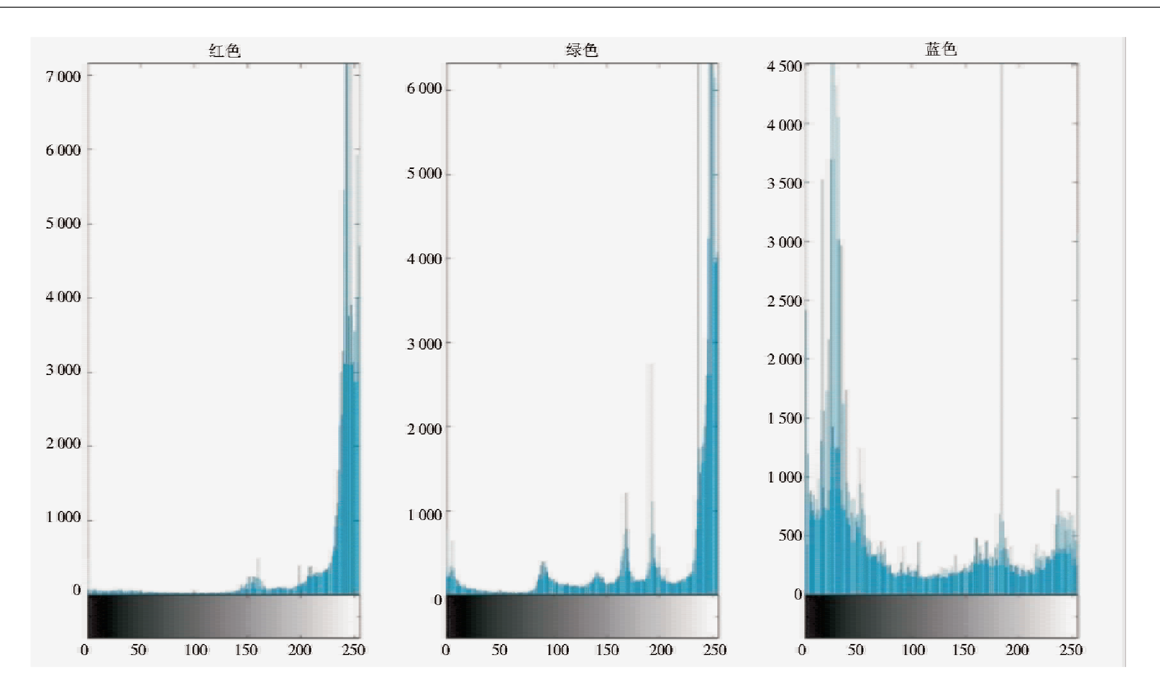


图 2　颜色直方图分量量化结果

够模拟人脑处理事务、识别图像、音频、视频，最终通过训练，让计算机获取如人脑一般地处理事务的能力。深度学习目前已广泛应用于人工智能领域，常见的模型有深度自动编码器(Deep Auto encoder，DAE)、深度受限玻尔兹曼机(Deep Restricted Boltzmann Machine，DRBM)以及卷积神经网络(Convolutional Neural Networks，CNN)，其中 CNN 模型在图像识别、人脸识别领域的运用已经取得较大成功[28]。

3.1　CNN 算法及其在图像情感抽取中的应用

卷积神经网络(Convolutional Neural Networks，CNN)已被证明可以从大量的数据系列中，学习出复杂的映射关系，在计算机视觉领域已经取得较大科研成果，被运用于图像分类、目标检测、表情识别、人脸检测等领域。相对于颜色特征、纹理特征的提取，深度卷积神经网络具有很强的抗干扰能力，对于图片的移动、旋转、形变或其他形式的变形具有高度的不敏感性[28]，其网络结构模型如图 3 所示：

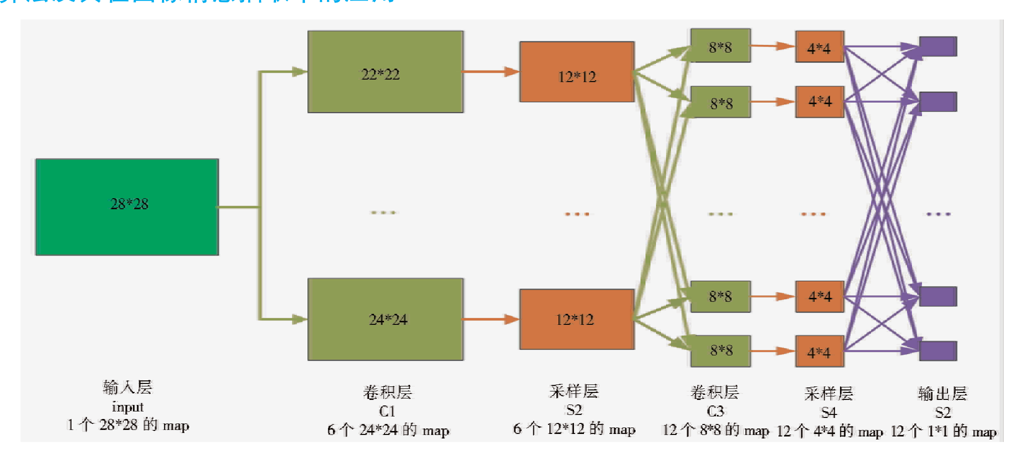


图 3　卷积神经网络模型结构

在 CNN 中，多层结构网络的可视层输入为每一个图像块。在卷积层中，通过一组滤波器以及非线性层变换，提取出目标图像的局部特征。输入的图像通过三个可训练的滤波器和可加偏置进行卷积，卷积后在 C1 层产生特征 map，然后对特征 map 中每组的四个像素进行求和，赋予权值、加入偏置，通过一个 Sigmoid 函数得到 6 个 S2 层的特征 map。具体计算过程如式(2)所示：

$$X_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right) \qquad 式(2)$$

其中，M_j 为输入该层的特征 map 集，l 表示当前的

卷积层数，X_j^l 表示 S 层第 j 个神经元输出的特征向量，x_i^{l-1} 为上一卷积层第 i 个神经元输出，k_{ij}^l 表示上一卷积层第 i 个神经元与本层第 j 个神经元之间的连接阈值。该过程所增加的偏置为 b_j^l，通过激活函数 f。

在 S2 层中，采用池化的方式对 C1 特征向量进行降维，以提升特征的表达能力。这些特征 map 再经过滤波得到 C3 层，该层级结构对每一个 map 重复如 S2 相同的计算过程，产生 S4。最终，这些像素值被光栅化，并连接成一个向量输入到传统的神经网络，得到输出[28]。具体计算公式如式(3)所示：

$$X_j^l = f\left[\left(\frac{1}{n}\sum_{i\in M_j} x_i^{l-1}\right) + b_j^l\right] \quad 式(3)$$

其中，n 表示从卷积层至采样层的窗口宽度。

为减少需要训练的参数数量，模型引入了优化方法：参数共享。观察发现，卷积层特征检测如垂直边缘检测如果适用于图片的某个区域，那么它也可能适用于图片的其他区域。如：采用一个 3×3 的过滤器检测垂直边缘，那么图片的左上角区域以及旁边的各个区域都可以使用这个 3×3 的过滤器。每个特征检测器以及输出都可以在输入图片的不同区域中使用同样的参数，以便提取垂直边缘或其它特征。

通过卷积层得到的特征向量维数通常较大，冗余信息很多，会增大训练代价，训练过程中易引起过拟合。因此需在 S 层的不同位置对特征向量进行降采样操作，即池化(POOL)。池化减少图像特征 map 输入的高度和宽度，它利于减少特征抽取计算量，并有助于使特征检测器的输入位置更加稳定。这两种池化层是：

(1)最大池化层：在输入图像特征上滑动(f,f)窗口，并将窗口的最大值存储在输出中。如图 4 所示，采用一个 2×2 的 filter(过滤器)滑动卷积层输出的图像特征，max pool 是在每一个区域中寻找最大值，这里的 stride=2，最终图像原特征经过不重叠提取的最大池化过程，得到图中右边的特征集，且原特征集参数降低，为原来的 1/4。

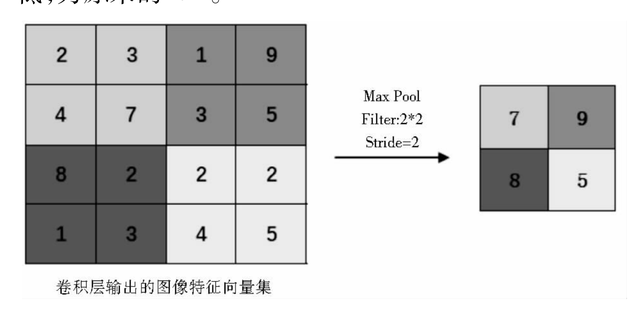


图 4 采样层最大池化操作

(2)平均池化层：在输入上滑动(f,f)窗口，并将窗口的平均值存储在输出中。如图 5 所示，同样采用一个 2×2 的 filter(过滤器)滑动卷积层输出的图像特征，avg pool 是在每一个领域特征点中求取平均值，更多地保留图像的背景特征。其中，stride=2，最终图像原特征经过不重叠提取的平均池化过程，得到图中右边新的特征集，进行后续训练，且原特征集参数降低，为原来的 1/4。

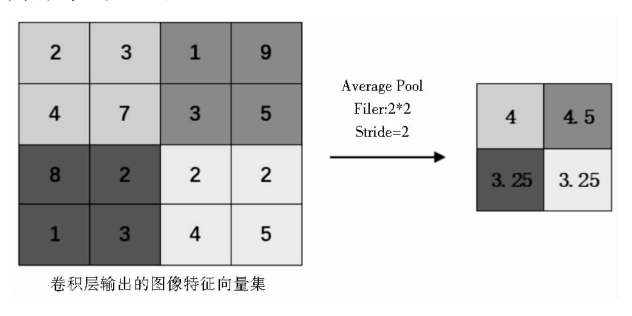


图 5 采样层平均池化操作

结合多层卷积层与池化层形成的深度网络模型能够取得对缩放、平移和旋转不变性的样本图片特征。

同时，卷积神经网络模型的训练过程采用反向传播算法，用于更新网络参数，如式(4)所示：

$$dA = \sum_{h=0}^{nH}\sum_{w=0}^{nH} W_c \times dZhw \quad 式(4)$$

式(4)中，W_c 是过滤器，$dZhw$ 是与第 h 行第 w 列的卷积层 Z 的输出梯度相对应的标量(对应于第 i 步向左和第 j 步向下抽取的点积)。在每次更新 dA 时，会将相同的滤波器 Wc 乘以不同的 dZ，该处理针对在计算正向传播时，每个过滤器都被不同的 a_{slice} 点平分和相加。因此，在计算 dA 的 backprop 时，只是添加所有 a_{slice} 的渐变。

卷积神经网络模型经过逐层分析抽取出图像特征，在卷积神经网络模型的最后一层引入损失函数，表示特征抽取的误检程度，引入损失函数 dWc(dWc 是一个滤波器的导数)，如式(5)所示：

$$dW_c = \sum_{h=0}^{nH}\sum_{w=0}^{nH} a_{slice} \times dZ_{hw} \quad 式(5)$$

式(5)中，a_{slice} 对应于切片，用于产生活性 Zij。计算结果产生与该切片相关的 W 渐变，由于 W 值相同，可将其相加获得 dW。

实验过程需训练出用于判断图像情感语义的卷积神经网络模型。不同的网络模型具有不同特性。在正式实验之前，选取不同的网络模型进行预训练、测试与反馈，最终选取测试效果较好的 CNN-VGGNet 模型作为实验的网络模型。在图像分类领域，牛津大学视觉几何组构建了 VGGNet-16 网络模型[29]，被证明是较为

李志义，许洪凯，段斌. 基于深度学习 CNN 模型的图像情感特征抽取研究[J]. 图书情报工作,2019,63(11):96－107.

优秀的图像特征抽取模型。但其缺陷也较为明显:一是该模型采用 6 段卷积层,由于层数较多,训练时产生较多的特征参数,对于图像情感特征训练参数数量需求较大,在多数情形下超过可供训练的样本数量不利于模型的收敛,不易达到最优值;二是网络共有 5 个最大池化层,较多的池化层易造成图像特征向量微粒化,使得图像主成分特征丢失,无法有效解决过拟合的问题。

3.2 CNN 模型的改进

由于在数据预训练中,情感特征抽取准确程度较低,过拟合情况明显,因此本实验改进了 VGGNet 模型和参数,重新进行了训练,以提升情感特征抽取效果。原卷积神经网络模型有三段卷积层和两个全连接层,并且针对不同特征修改卷积层中滤波器的大小和数量。根据样本数量以及预训练结果,对网络模型进行改进,并对原模型参数进行微调:①采用 5 层卷积层(conv2d),该改进方法在原模型的基础上,针对图片的不变性所做出的优化,有利于防止过拟合,提高模型的泛化能力。②设置 4 层最大池化层(maxpool);③将连接方式改进为 3 层全连接(Dense);④为减少过拟合的影响,放弃(Dropout)一部分神经元;改进后模型的参数如表 1 所示:

表 1 改进的 VGGNet 模型参数

Layer (type)	Output	Shape	Param #
conv2d_2S (Conv2D)	(None,	32, 32, 64)	4864
max_pooling2d_22 (MaxPooling	(None,	16, 16, 64)	0
conv2d_30 (Conv2 D)	(None,	16, 16, 64)	36928
dropout_43 (Dropout)	(None,	16, 16, 64)	0
conv2d_31 (Conv2D)	(None,	16, 16, 64)	36928
dropout_44 (Dropout)	(None,	16, 16, 64)	0
conv2d_32 (Conv2 D)	(None,	16, 16, 32)	18464
max_pooling2d_23 (MaxPooling	(None,	8, 8, 32)	0
conv2d_33 (Conv2 D)	(None,	8, 8, 64)	18496
dropout_45 (Dropout)	(None,	8, 8, 64)	0
conv2d_34 (Conv2D)	(None,	8, 8, 64)	36928
max_pooling2d_24 (MaxPooling	(None,	4, 4, 64)	0
conv2d_35 (Conv2D)	(None,	4, 4, 128)	73856
dropout_46 (Dropout)	(None,	4, 4, 128)	0
conv2d_36 (Conv2D)	(None,	4, 4, 128)	147584
max_pooling2d_25 (MaxPooling	(None,	2, 2, 128)	0
flatten_ll (Flatten)	(None,	512)	0
dropout_47 (Dropout)	(None,	512)	0
dense_21 (Dense)	(None,	2500)	1282500
dropout_48 (Dropout)	(None,	2500)	0
dense_22 (Dense)	(None,	1024)	2561024
dropout_49 (Dropout)	(None,	1024)	0
dense_23 (Dense)	(None,	2)	2050
local params: 4,219,622			
Trainable params: 4,219,622			
Non-trainable params: 0			
None			

改进后模型能够以较少的参数与迭代次数完成图像情感特征抽取,所加深的卷积层对图像情感特征有更高的敏感性。同时,四层采样层均采用最大池化的方式对情感特征向量进行降维,有效地增加了图像情感特征的显著性,对于数据集具有一定的泛化能力。

改进的 CNN 模型算法用代码可描述为:

```
# 用来创建卷积层并把本层的参数存入参数列表
# input_op:输入的 tensor name:该层的名称 kernel:卷积层 n_out:输出通道数,dh:步长的高 dw:步长的宽,p 是参数列表
# 定义卷积层
def conv_op(input_op,name,kh,kw,n_out,dh,dw,p):
    conv = tf.nn.conv2d(input_op, kernel, (1,dh,dw,1),padding='SAME')
    activation = tf.nn.relu(conv,name=scope)
    return activation
# 定义全连接层
def fc_op(input_op,name,n_out,p):
    activation = tf.nn.relu_layer(input_op,kernel,biases,name=scope)
    return activation
# 定义最大池化层
def mpool_op(input_op,name,kh,kw,dh,dw):
    return max_pool(input_op,ksize=[1,kh,kw,1],strides=[1,dh,dw,1],padding='SAME',name=name)
#定义网络结构
def inference_op(input_op,keep_prob):
    p = []
    conv1_1 = conv_op(input_op,name='conv1_1',kh=3,kw=3,n_out=32,dh=1,dw=1,p=p)
    pool1 = mpool_op(conv1_2,name='pool1',kh=2,kw=2,dw=2,dh=2)
    conv2_1 = conv_op(pool1,name='conv2_1',kh=3,kw=3,n_out=64,dh=1,dw=1,p=p)
    dropout2_1 = dropout(conv2_1, keep_prob, name='dropout2_1')
    conv2_2 = conv_op(dropout2_1,name='conv2_2',kh=3,kw=3,n_out=64,dh=1,dw=1,p=p)
    dropout2_2 = dropout(conv2_2,keep_prob, name='dropout2_2')
    conv2_3 = conv_op(dropout2_2,name='conv2_3',kh=3,kw=3,n_out=32,dh=1,dw=1,p=p)    pool2 = mpool_op(conv2_3, name='pool2', kh=2, kw=2, dw=2, dh=2)
    conv3_1 = conv_op(pool2, name='conv3_1', kh=3, kw=3, n_out=64, dh=1, dw=1, p=p)
    dropout3_1 = dropout(conv3_1, keep_prob, name='dropout3_1')
    conv3_2 = conv_op(dropout3_1, name='conv3_2', kh=3, kw=3, n_out=64, dh=1, dw=1, p=p)
    pool3 = mpool_op(conv3_2, name='pool3', kh=2, kw=2, dw=2, dh=2)
    conv4_1 = conv_op(pool3, name='conv4_1', kh=3, kw=3, n_
```

```
out = 128, dh = 1, dw = 1, p = p)
    dropout4_1 = dropout(conv4_1, keep_prob, name = 'dropout3_1')
    conv4_2 = conv_op(dropout4_1, name = 'conv4_2', kh = 3, kw = 3,
n_out = 128, dh = 1, dw = 1, p = p)
    pool4 = mpool_op(conv4_2, name = 'pool4', kh = 2, kw = 2, dw =
2, dh = 2)
    #平铺层
    shp = pool5.get_shape()
    flattened_shape = shp[1].value * shp[2].value * shp[3].value
    resh1 = tf.reshape(pool5,[ -1,flattened_shape],name = "resh1")
    fc5_drop = dropout(resh1, keep_prob,name = "fc5_drop")
    fc6 = fc_op(resh1,name = "fc5_drop",n_out = 2500,p = p)
    fc6_drop = dropout(fc6,keep_prob,name = 'fc6_drop')
    fc7 = fc_op(fc6_drop,name = "fc7",n_out = 2500,p = p)
    fc7_drop = tf.nn.dropout(fc7,keep_prob,name = "fc7_drop")
    fc8 = fc_op(fc7_drop,name = "fc8",n_out = 2,p = p)
    softmax = tf.nn.softmax(fc8)
    predictions = tf.argmax(softmax,1)
    return predictions,softmax,fc8,p
```

上述代码引入 6 层循环,卷积层收到 N 个 32 ×32 ×3 特征图作为输入,经过宽度为 K × K 的卷积核,产生一个输出特征图的像素点。卷积过程所引入的滑动窗 stride 间隔为 S,小于 K,产生的 M 个特征输出图用于下一层卷积层。

3.3 构建图像情感抽取实验模型

实验过程需训练出用于判断图像情感语义的卷积神经网络模型。依据改进后模型的参数、结构,结合本文实验目的,构建出改进后的 CNN 网络模型,如图 7 所示。为确保实验过程的科学性与灵活性,探究网络深度对模型精确度的影响,所有卷积层采用相同的配置:①卷积核大小为 3 ×3,步长为 1(stride:1),填充为 1(padding:1);②共有 4 个最大池化层,大小均为 2 × 2,步长为 2(stride:2);③三个全连接层,前两层都有 4 096通道,第三层共 1 000 路即代表 1 000 个标签类别,最后一层为 softmax 层,完成图像情感特征抽取分类;④所有隐藏层后都带有 ReLU 非线性激活函数。如图 6 所示:该 VGGNet 模型共 16 层,13 层卷积层与 3 层全连接层。首层为卷积层,将输入图像的大小变换为 32 ×32 ×64,在 3 个通道上采用 3 ×3 的滤波器对图像进行卷积操作后,得到基本的卷积数据;然后进行一次 ReLU 和 Norm 变换,随后通过采样层进行最大池化,得到规模为 16 ×16 ×32 的特征图作为输出传入下一层。第 2 层也是卷积层,处理过程和第 1 层类似,经过卷积、ReLU、Norm 和池化,输出结果规模为 8 ×8 ×64,卷积层处理过程基本相同。第 14 层进入全连接层,全连接后特征向量达到 4 096 个节点,最后一层全连接层是对上一个全连接层进行 ReLU 以及 dropout 后再进行全连接的结果,最后输出为融合 label 的 softmax loss。

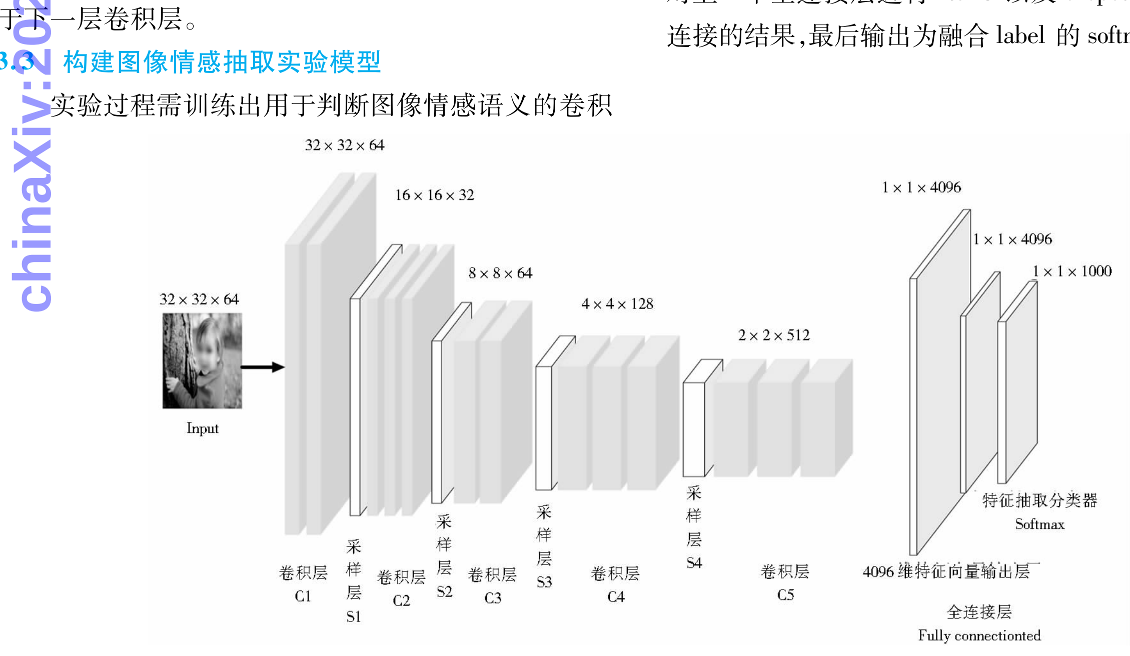


图 6 改进的卷积神经网络模型 VGGNet-16

4 实验分析与评价

图像情感特征提取主要采用深度模型下的小样本实验方法,训练出能够抽取图像情感特征的深度学习模型。

4.1 数据来源

由于 Flickr 网站具备了较为成熟的情感图像数据集[30],本实验基于有监督学习的方法,精选其中5 000 张图像,形成 FlickrEmotion 数据集。其图像均来源于社交媒体上用户所上传的图像,图像规格较高并且可

以完成高清无损转换。该数据集图像训练语料具备情感标签(Label),在分类标签下共分为 2 个情感类:一类为积极情感,包含开心、高兴、愉悦;另一类为消极情感,包含伤心、沮丧、压抑。为进行图像情感特征抽取研究,本文将此 6 种标签进行进一步合并为仅具有积极和消极两种情感,即二分类情感特征,经 Softmax 分类器验证,图像情感均符合标签,分类有效。最后,将该数据集分为训练样本和测试样本,有 2 500 张训练样本和 2 500 张测试样本。图 7、图 8 展示了所选图像数据集的部分实例。



图 7　积极类情感图像

图像来源:http://press. liacs. nl/mirflickr/mirflickr25k/mirflickr25k_annotations_v070. zip



图 8　消极类情感图像

图像来源:http://press. liacs. nl/mirflickr/mirflickr25k/mirflickr25k_annotations_v070. zip

4.2　数据预处理

为使卷积神经网络更好地采集数据集中图像的特征,需对图片进行批量预处理。由于图像的颜色特征与纹理特征能够有效地表达图片的情感,因此可先通过预处理对图像纹理特征与色彩特征进行抽取,然后通过改变二者权重输入卷积神经网络模型。

如图 9 所示,深化纹理特征,淡化色彩特征,得到能够被 VGGNet 识别的边界像素,数据处理的结果使图像纹理特征增强。

如图 10 所示,将图片依照图像颜色特征抽取方法进行预处理,即深化图像的颜色特征,将像素降低,减小边缘特征的比重。由于像素的降低,使得图像局部颜色的比重升高,达到了深化色彩特征、淡化纹理特征的效果。

将预处理图像输入 CNN 模型,以测试实验模型情感特征抽取效果。在不改变学习率(0.01)、卷积层滤波器大小为 3×3 的条件下,经过训练,初步得到情感特征抽取结果。如表 2 所示,经过两类图像增强处理,预训练集的图像情感特征抽取准确率能达到较好的水平,均维持在 78% 左右,达到了较高的准确率。但特征抽取过程存在检索损失,即开始出现检索出无关特征的情况,出现了过拟合。预处理过程无差错,表明改进后 VGGNet 模型已具有较好的鲁棒性,可以将其运用于实验过程,要提升准确率只需进行微调操作。

表 2　预训练集图像情感特征抽取准确率

训练集预处理类型	维度	准确率	训练结果评价
减少颜色特征,增加边缘特征	4096	77.35%	过拟合
增加颜色特征,减少边缘特征	4096	78.33%	过拟合

4.3　图像特征抽取实验与分析

为确保实验过程的科学性与灵活性,探究网络深度对模型精确度的影响,所有卷积层采用相同的配置,并且以本实验模型为基础,根据数据预处理结果,相应地对情感特征抽取过程做如下改变:

(1)对于加强纹理特征的网络,滤波器大小调整

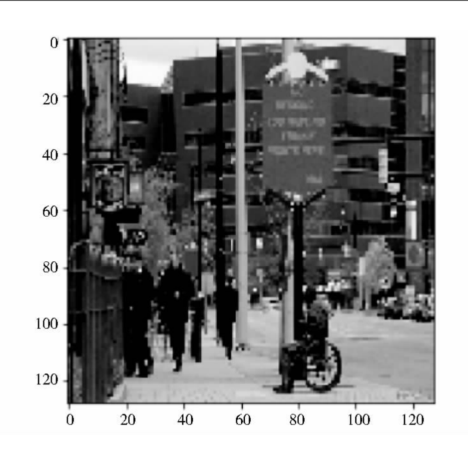


图9 减少颜色特征、增加边缘特征预处理

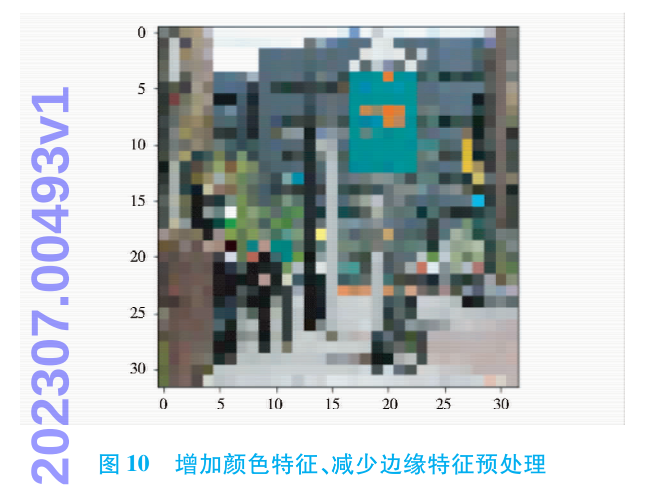


图10 增加颜色特征、减少边缘特征预处理

为5×5。

(2)对于加强色彩特征的网络,滤波器大小调整为3×3。

为了提高准确率,实验过程尝试继续增加卷积层,但出现过拟的情况;试图降低学习率,导致网络无法收敛;提高学习率,使部分神经元出现死亡的情况,因此将学习率调至0.1。

使用训练好的模型对图片进行情感特征提取:训练结果越趋近于1,其积极情感愈强烈;反之则偏向于负面情绪。测试结果如图11所示,将图像颜色特征增强,并以此输入卷积网络模型,该样例为积极情感的概率为85.22%,符合测试集label。

如图12所示,强化图像纹理特征,降低颜色空间,通过提取的纹理特征,判断出图像的情感值,样例为积极情感的概率为84.32%,符合label。

训练结果表明,通过改进的卷积神经网络抽取图像颜色特征或纹理特征,判断出图像的情感语义,能取得较好的结果。相对于利用原VGGNet16模型所进行的情感特征抽取(见表1),改进后的模型对情感语义的判断能力提升10个百分点。可见,本次算法的改进

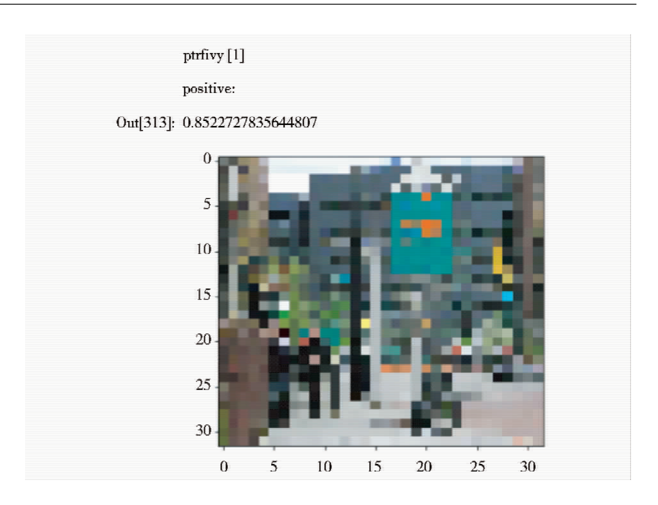


图11 图像训练结果及其准确率

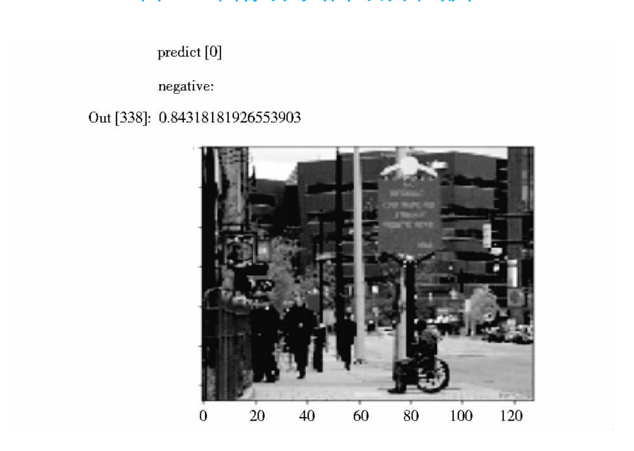


图12 图像训练结果及其准确率

能够有效地提升图像情感特征抽取的准确率。

将改进后的模型运用到数据集中进行训练,并验证其性能。对图13观测可知,在0至50次迭代中,训练集与测试集的准确率逐渐上升;期间测试集准确率达到峰值,出现在第20到30次迭代中,其准确率达到65%;并能较为稳定地描述情感特征,在60次后出现过拟合,随后训练集准确率与测试集准确率出现分离趋势。第50次至100次迭代中,训练集准确率呈上升趋势,而测试集准确率呈下降趋势,并在57%准确率附近稳定波动。这反映了深度网络模型训练中出现过拟合的情况。其主要原因为数据样本较小,训练误差很小,而泛化误差很大,同时反映了使用边缘特征分辨图像感情的不足。

基于颜色增强的情感特征抽取结果如图14所示:在0至80次迭代中,训练集与测试集准确率逐渐上升;期间测试集准确率达到峰值(76%)是出现在第60次到79次迭代中;此外,测试集有较高的准确率(75%),并且在情感特征的获取方面较为稳定,过拟合情况不明显。随后训练集准确率与测试集准确率出

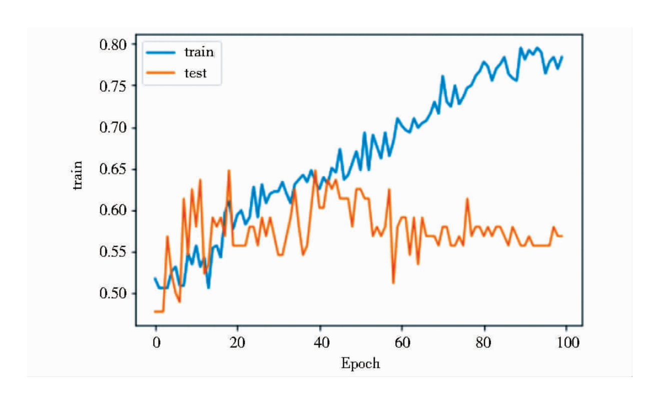


图 13　基于纹理增强的情感特征抽取结果

现逐渐分离趋势，在第 80 次至 100 次迭代中，训练集准确率呈上升趋势，而测试集准确率在 65% 附近稳定波动，有过拟合的趋势。

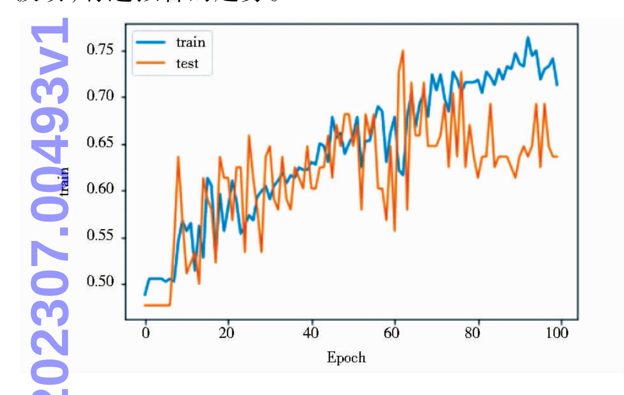


图 14　基于颜色增强的情感特征抽取结果

比较图 13 和图 14 两组以不同方式进行数据增强的图像情感特征抽取结果发现：在样本数量相同，深度网络模型一致的情况下，使用经过减少颜色特征，增加边缘特征处理后的数据训练出的模型测试集，其准确率峰值更高，稳定值也更高。由此进一步得知，图像底层颜色、纹理特征，能够与高层情感语义建立映射关系，并且在这个关系中，颜色特征对情感特征抽取展现出更好的感知力。

为直观地观测实验结果，将两种不同的图像处理类型的情感特征抽取结果进行了列表，如表 3 所示，从维度、准确率和训练结果评价三方面进行了比较。

实验证明：基于深度学习的图像情感特征抽取，能够有效地判定图像的情感倾向，通过该算法，能够摆脱图像情感特征抽取的人工干预，充分利用图像本身的属性进行图像情感的判定，并且达到了较为满意的准确率。这种基于图像情感语义特征抽取的检索方法适应性强，可应用于多种领域：如人脸识别的情感计算，通过抽取人像面部的表情特征判断其所处的情感状态[31]；通过分析驾驶员的面部情感特征，判断驾驶过程的安全性等。

5　结语

本文是对图像情感语义特征抽取的有益探索，在分析卷积神经网络模型与传统的图像特征抽取方法的基础上，使用了改进的 VGGNet 模型抽取图像的情感特征，在图像数据集上取得了 80.21% 的检准率，相较于原模型，改进后的模型在检准率上提升了 10%，达到了实验的目的。此外，通过优化网络模型的参数设置，有助于更有效地抽取图像情感特征，获得了较好的情感特征抽取结果。

表 3　改进的 VGGNet 图像情感检索模型的准确率对比

数据集 / 图像处理类型	测试集			训练集		
	维度	准确率	训练结果评价	维度	准确率	训练结果评价
减少颜色特征，增加边缘特征	4 096	65.35%	较低	4 096	80.21%	有效
增加颜色特征，减少边缘特征	4 096	75.2%	有效	4 096	77.4%	有效

跨模态检索是未来信息检索领域的发展方向，以文字检索图片是跨模态领域最基本的检索方式，而基于图像情感语义的检索则是该领域前沿性较强的一个课题[32－34]。目前，研究人员已经将情感语义检索运用于图像分类、人脸识别、表情判断等诸多领域[35]。本文由于实验过程所搭建的开发平台受限于机器配置，在卷积网络模型中特征向量的输出维度仅达到 4096 维。但通过加深网络，也能达到预期的训练结果。

参考文献：

[1] 曾金，陆娜，胡潇戈等. 网站新闻人物图像情感倾向研究[J]. 情报科学,2018,36(6):131－135,141.
[2] 马松岳，许鑫. 基于评论情感分析的用户在线评价研究——以豆瓣网电影为例[J]. 图书情报工作,2016,60(10):95－102.
[3] 蒋知义，马王荣，邹凯等. 基于情感倾向性分析的网络舆情情感演化特征研究[J]. 现代情报,2018,38(4):50－57.
[4] 黄崑，赖茂生. 以用户情感为线索的图像检索研究[J]. 情报科学,2006(09):1395－1399.
[5] 汤丽萍，陈芬. 基于情感的图像分类研究进展[J]. 情报理论与实践,2018,41(6):149－153,160.
[6] 李志义，黄子风，许晓棉. 基于表示学习的跨模态检索模型与特征抽取研究综述[J]. 情报学报,2018,37(4):422－435.
[7] 曾金，陆伟，陈海华等. 基于多模数据的微博用户兴趣识别研究[J]. 情报科学,2018,36(1):124－129.

[8] BLACK J A, KUCHI P, PANCHANATHAN S. Indexing natural images for retrieval based on Kansei factors[J]. Proceedings of SPIE -The International Society for Optical Engineering, 2004, 5292:363 –375.

[9] YOSHIDA K, KATO T, YANARU T. Image retrieval system using impression words[C]// IEEE international conference on systems, man, and cybernetics. IEEE, 1998:2780 –2784 vol. 3.

[10] CHO S B, LEE J Y. A human-oriented image retrieval system using interactive genetic algorithm[J]. Systems man& cybernetics Part A Systems & humans IEEE Transactions on, 2002, 32(3):452 –458.

[11] COLOMBO C, DEL BIMBO A, Pala P. Semantics in visual information retrieval[J]. Multimedia IEEE, 1999, 6(3):38 –53.

[12] SIERSDORFER S, MINACK E, DENG F, et al. Analyzing and predicting sentiment of images on the social Web[C]//Proc of international conference on multimedia. New York: ACM Press, 2010:715 –718.

[13] YOU Q, LUO J, JIN H, et al. Robust image sentiment analysis using progressively trained and domain transferred deep networks [J]. American Association for Artificial Intelligence, 2015:381 –388.

[14] ORIOL V, ALEXANDER T, SAMY B, et al. Show and tell: A neural image caption generator[C]//Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2015:3156 –3164.

[15] SUN M, SONG Z, JIANG X, et al. Learning pooling for convolutional neural network[J]. Neurocomputing, 2017, 224:96 –104.

[16] 黄崑,骆方,游祎. 图像情感特征及其检索应用[J]. 情报科学, 2010, 28(04):602 –606.

[17] 王上飞,薛佳,王煦法. 基于内容的情感图像获取模型[J]. 计算机科学, 2004(9):186 –19.

[18] 武频,陶聪,朱永华等. 基于情感语义的图像注释与检索[J]. 计算机技术与发展, 2015, 25(10):13 –18.

[19] 王华秋,胡立松. 图像情感语义分类及检索研究[J]. 重庆理工大学学报(自然科学), 2017, 31(10):180 –186.

[20] 谭莲芝. 基于深度学习的人脸属性识别方法研究[D]. 北京:中国科学院大学(中国科学院深圳先进技术研究院), 2017.

[21] 陈金菊,欧石燕. 数字图像语义标注模型比较与分析[J]. 图书情报工作, 2018, 62(6):116 –124.

[22] 陈静,陈德照,陆泉. 一种基于表情分析的图像用户兴趣提取模型[J]. 情报理论与实践, 2014, 37(5):92 –96, 106.

[23] 姜亚茜. 考虑情感因素的图像颜色编辑[D]. 天津:天津大学, 2017.

[24] 徐彤阳,任浩然,张国标. 我国图像检索领域的前沿动态研究探析——CNKI 数据收录文献计量分析[J]. 图书馆学研究, 2017(4):2 –10.

[25] 成琳,陈俊杰,相洁. 图像颜色征提取技术的研究与应用[J]. 计算机工程与设计, 2009, 30(14):3451 –3454.

[26] 王彦林. 基于 HSV 颜色特征图像检索算法在 Matlab 中的实现[J]. 电脑编程技巧与维护, 2013(16):86 –87.

[27] 杨瑞. 基于颜色情感语义的检索系统的设计与实现[D]. 开封:河南大学, 2015.

[28] 张军阳,王慧丽,郭阳,等. 深度学习相关研究综述[J]. 计算机应用研究, 2018, 35(07):1921 –1928, 1936.

[29] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. Computer science, 2014. arXiv:1409.1556.

[30] 陆伟,罗梦奇,丁恒等. 深度学习图像标注与用户标注比较研究[J]. 数据分析与知识发现, 2018, 2(5):1 –10.

[31] 朱虹,李千目,李德强. 基于单个卷积神经网络的面部多特征点定位[J]. 计算机科学, 2018, 45(4):273 –277, 284.

[32] 胡二雷,冯瑞. 基于深度学习的图像检索系统[J]. 计算机系统应用, 2017, 26(3):8 –19.

[33] 任夏荔,陈光喜,曹建收等. 基于深度学习特征的图像检索方法[J]. 计算机工程与设计, 2018, 39(2):503 –510.

[34] 吴琼. 基于卷积神经网络的图文情感分析技术研究[D]. 泉州:华侨大学, 2017.

[35] 金汉均,段贝贝. 卷积神经网络在跨媒体检索中的应用研究[J]. 电子测量技术, 2018, 41(7):54 –57.

作者贡献说明:

李志义:拟定题目、提出研究思路,算法代码撰写,论文修改与定稿;

许洪凯:论文撰写、算法改进、检验、参考文献分析;

段斌:平台搭建、算法改进、参数设置。

Research on Image Emotion Feature Extraction Based on Deep Learning CNN Model

Li Zhiyi[1] Xu Hongkai[1] Duan Bin[2]

[1] Economic & Management College of South China Normal University, Guangzhou 510006

[2] Information & Photoelectric Science College of South China Normal University, Guangzhou 510006

Abstract: [**Purpose/significance**] Image retrieval based on user emotion has become a hotspot in machine learning research. However, the corpus data of image sentiment feature annotation is mostly derived from the extraction of low-level features of images, which leads to the simplification and stylization of image retrieval process. The algorithm of image emotion feature extraction based on deep learning fuses the underlying features of the image into the high-level emotion seman-

tics of the image, which provides a reference for the emotional semantic retrieval of images. [**Method/process**] Using the improved convolutional network model, the color and texture of the dataset image were taken as input, the emotion information of the image was automatically extracted by multi-layer operation, and the sentiment retrieval accuracy of the improved model was calculated though the back propagation algorithm, and an image sentiment feature extraction model with high rate and low degree of over-fitting was constructed. [**Result/conclusion**] This paper completes the extraction of emotional features of the image through an improved deep convolutional network model, and improves the retrieval accuracy by 10%.

Keywords: deep learning　image　emotional features　extraction　convolutional neural network

2018－2022 年教育部高等学校图书馆学专业教学指导委员会成立会暨第一次工作会议纪要

2019 年 1 月 18 日,2018－2022 年教育部高等学校图书馆学专业教学指导委员会(以下简称“教指委”)成立会暨第一次工作会议在上海大学举行。2018－2022 年教育部高等学校图书馆学专业教指委主任委员、北京大学信息管理系王余光教授,副主任委员武汉大学研究生院院长陈传夫教授、南京大学郑建明教授、华中师范大学副校长夏立新教授、上海大学图书情报档案系吕斌教授,秘书长北京大学信息管理系党委书记张久珍教授,以及来自国内各图书馆学专业教学单位的 30 余名委员,齐聚上海大学,庆祝本届教指委的成立,并对未来五年的工作进行规划安排。上海大学党委常委、副校长聂清出席会议并致辞,上海大学教务处处长彭章友教授,上海大学图书情报档案系主任金波教授、系党委书记丁华东教授出席开幕式,部分专业教师代表列席会议。

成立大会由图书馆学专业教学指导委员会副主任吕斌主持。上海大学聂清副校长在致辞中对本届图书馆学教指委成立会于上海大学召开表示热烈欢迎,向与会专家介绍了上海大学创建高水平大学的概况和本科教育教学方面的主要举措,并对本次会议的顺利举行表达了衷心祝愿。教育部图书馆学专业教指委主任委员王余光教授代表本届教指委向上海大学及会议具体承办单位上海大学图书情报档案系表达了感谢,向各位委员传达了教育部高等学校专业教学指导委员会会议讲话精神,并代表教育部向 2018－2022 年教育部高等学校图书馆学专业教学指导委员会成员颁发了聘书。

开幕式结束后,本届教指委第一次工作会议随即举行,会议由夏立新副主任主持。陈传夫副主任代表上一届教指委做了《上一届教育部高等学校图书馆学专业教学指导委员会工作总结》,对图书馆学教指委的发展历程及前三届教指委所做的主要工作进行了回顾总结。总结报告结束后,按照教育部要求,陈传夫副主任带领全体委员共同学习了《教育部高等学校教学指导委员会章程》,以及由上一届图书馆学教指委制定的《图书馆学本科专业教学质量国家标准》。王余光主任对上午会议的相关情况进行了总结,并结合前三届教指委工作经验,对本届教指委的主要职责进行了说明。

下午会议由郑建明副主任主持。张久珍秘书长向全体委员传达了教育部高等学校教学指导委员会成立大会的会议精神,要求本届教指委要切实贯彻教育部“以本为本”“全面振兴本科教育”的要求,为“建设一流本科、做强一流专业、培养一流人才”做出贡献。随后她代表本届教指委汇报《2018－2022 年教育部高等学校图书馆学专业教学指导委员会工作计划》(以下简称《工作计划》),包括国标的贯彻、宣传和培训,提升图书馆学教学质量,图书馆学课程建设,图书馆学教材建设,图书馆学人才培养研究和创新型人才培养,图书馆学师资建设等方面。《工作计划》经委员会讨论通过,明确了教指委下一阶段的工作方向。

吕斌副主任主持了随后的研讨环节,武汉大学研究生院院长陈传夫副主任、华中师范大学副校长夏立新副主任、兰州大学副校长沙勇忠委员、中国科学院大学初景利委员、南京农业大学人文社科处处长黄水清委员、复旦大学图书馆副馆长杨光辉委员、中国人民大学索传军委员、东北师范大学王战林委员等相继发言,从不同角度对本届教指委工作提出了意见和建议。会议讨论的成果,将在整理后报送教育部批准实施。

王余光主任最后对本次会议进行了全面总结,对 2019 年第二次工作会议做出了安排。会议在热烈的气氛中完成各项议程,圆满闭幕。

(2018－2022 年教育部高等学校图书馆学专业教学指导委员会委员名单见第 126 页)

(张久珍)